

Hypotheses tests in boundary regression models*

HOLGER DREES, NATALIE NEUMEYER AND LEONIE SELK

University of Hamburg, Department of Mathematics

Bundesstrasse 55, 20146 Hamburg, Germany

August 19, 2014

Abstract

We consider a nonparametric regression model with one-sided errors and regression function in a general Hölder class. Our aim is inference on the error distribution. To this end we estimate the regression function via minimization of the local integral of a polynomial approximation. We show uniform rates of convergence for the simple regression estimator as well as for a smooth version. Those rates continue to hold in mean regression models with symmetric and bounded error distribution. Here one obtains faster rates of convergence compared to optimal rates in mean regression when the error distribution is irregular in the sense that sufficient mass is concentrated near the endpoints. The results are applied to prove asymptotic \sqrt{n} -equivalence of a residual-based empirical distribution function to the empirical distribution function of unobserved errors in the case of irregular error distributions. This result is remarkably different from corresponding results in mean regression with regular errors. It can readily be applied to develop goodness-of-fit tests for the error distribution. We present some examples and investigate the small sample performance in a simulation study.

AMS 2010 Classification: Primary 62G08; Secondary 62G10, 62G30, 62G32

Keywords and Phrases: goodness-of-fit testing, irregular error distribution, one-sided errors, residual empirical distribution function, uniform rates of convergence

1 Introduction

Boundary regression models arise naturally in image analysis, analysis of auctions and records, or in extreme value analysis with covariates. We consider such a model where

*Financial support by the DFG (Research Unit FOR 1735 *Structural Inference in Statistics: Adaptation and Efficiency*) is gratefully acknowledged.

the boundary regression function g is defined as the right endpoint of the conditional distribution of the response Y , given the covariate X . For such a boundary regression model with multivariate random covariates and twice differentiable regression functions, Hall and Van Keilegom (2009) establish a minimax rate for estimation of $g(x)$ (for fixed x) under quadratic loss and determined pointwise asymptotic distributions of an estimator which is defined as a solution of a linear optimization problem (cf. Remark 2.5). Müller and Wefelmeyer (2010) consider a mean regression model with (unknown) symmetric support of the error distribution and Hölder continuous regression function. They discuss pointwise MSE rates for estimators of the regression function that are defined as the average of local maxima and local minima. Meister and Reiß (2013) consider a regression model with known bounded support of the errors. They show asymptotic equivalence in the strong LeCam sense to a continuous-time Poisson point process model when the error density has a jump at the endpoint of its support. For a regression model with error distribution that is one-sided and regularly varying at 0 with index $\alpha > 0$, Jirak, Meister and Reiß (2014) suggest an estimator for the boundary regression function which adapts simultaneously to the unknown smoothness of the regression function and to the unknown extreme value index α . Reiß and Selk (2014) construct efficient and unbiased estimators of linear functionals of the regression function in the case of exponentially distributed errors as well as in the limiting Poisson point process experiment by Meister and Reiß (2013).

Closely related to regression estimation in models with one-sided errors is the estimation of a boundary function g based on a sample from (X, Y) with support $\{(x, y) \in [0, 1] \times [0, \infty] \mid y \leq g(x)\}$. For such models Härdle et al. (1995) and Hall et al. (1998) proved minimax rates both for $g(x)$ and for the L_1 -distance between g and its estimator. Moreover, they showed that an approach using local polynomial approximations of g yields this optimal rate. Explicit estimators in terms of higher order moments were proposed and analyzed by Girard and Jacob (2008) and Girard et al. (2013). In this setting, there is also an extensive literature on the estimation of monotonically increasing boundary functions g , which naturally arise in production frontier models; see Gijbels et al. (1999) and the literature cited therein.

The aim of the paper is inference on the error distribution in regression models with one-sided errors. To this end we need uniform rates of convergence for the regression estimator. To obtain those rates for general smoothness order $\beta \in (0, \infty)$ of the regression function and extreme value index $\alpha \in (0, \infty)$ of the error distribution is an end of its own. To the authors' knowledge no such uniform rates are available so far in the literature. The results can also be applied to mean regression models with bounded symmetric error distribution. For regression functions g in a Hölder class of order β we obtain the rate $((\log n)/n)^{\beta/(\alpha\beta+1)}$, where α denotes the extreme value index of the error distribution. Thus for $\alpha \in (0, 2)$ the rate is faster than the typical rate one has in mean regression models with regular errors. For pointwise and L^p -rates of convergence it has been known in the literature that faster rates are possible for nonparametric regression estimation in models with irregular error

distribution, see e. g. Gijbels and Peng (2000), Hall and Van Keilegom (2009), or Müller and Wefelmeyer (2010).

The uniform rate of convergence for the regression estimator enables us to derive asymptotic expansions for residual-based empirical distribution functions and weak convergence of the residual-based empirical distribution function. We state conditions under which the influence of the regression estimation is negligible such that the same results are obtained as in the iid-case (i. e. in the case of observable errors). We apply those results to derive goodness-of-fit tests for parametric classes of error distributions. Asymptotic properties of residual empirical distribution functions in mean regression models with regular errors were investigated by Akritas and Van Keilegom (2001), among others. Here in contrast to our results the regression estimation vastly influences the asymptotic behavior of the empirical distribution function. As a consequence asymptotic distributions of goodness-of-fit test statistics are complicated and typically bootstrap is applied, see Neumeyer, Dette and Nagel (2006).

The remainder of the article is organized as follows. In section 2 the regression model under consideration is presented and model assumptions are formulated. The regression estimator is defined and uniform rates of convergence are given. A smooth modification of the estimator is considered and uniform rates of convergence for this estimator as well as its derivative are shown. In section 3 residual based empirical distribution functions based on both regression estimators are investigated. Conditions are stated under which the influence of regression estimation is asymptotically \sqrt{n} -negligible. Further an expansion of the residual empirical distribution function is shown that is valid under more general conditions. Goodness-of-fit tests are discussed in general and in some detailed examples. A small simulation study shows the small sample performance of the tests. All proofs are given in the appendix.

2 The regression function: uniform rates of convergence

We consider a regression model with fixed equidistant design and one-sided errors,

$$Y_i = g\left(\frac{i}{n}\right) + \varepsilon_i, \quad i = 1, \dots, n, \quad (2.1)$$

under the following assumptions.

(F1) The errors $\varepsilon_1, \dots, \varepsilon_n$ are independent and identically distributed and supported on $(-\infty, 0]$. The error distribution function fulfills

$$F(y) = 1 - c|y|^\alpha + r(y)$$

for some $\alpha > 0$, with $r(y) = o(|y|^\alpha)$ for $y \nearrow 0$.

(G1) The regression function g belongs to some Hölder class of order $\beta \in (0, \infty)$, i.e. g is $\lfloor \beta \rfloor$ -times differentiable on $[0, 1]$ and the $\lfloor \beta \rfloor$ -th derivative satisfies

$$c_g := \sup_{\substack{t, x \in [0, 1] \\ t \neq x}} \frac{|g^{(\lfloor \beta \rfloor)}(t) - g^{(\lfloor \beta \rfloor)}(x)|}{|t - x|^{\beta - \lfloor \beta \rfloor}} < \infty.$$

For the estimator of the regression function we need the following assumption.

(H1) Let $(h_n)_{n \in \mathbb{N}}$ be a sequence of positive bandwidths that satisfies $\lim_{n \rightarrow \infty} h_n = 0$ and $\lim_{n \rightarrow \infty} n h_n |\log h_n|^{-1} = \infty$.

We consider an estimator that locally approximates the regression function by a polynomial while lying above the data points. More specifically, for $x \in [0, 1]$ let

$$\hat{g}_n(x) := \hat{g}(x) := p(x)$$

where p is a polynomial of order $\lceil \beta \rceil - 1$ and minimizes the local integral

$$\int_{x-h_n}^{x+h_n} p(t) dt \quad (2.2)$$

under the constraints $p(\frac{j}{n}) \geq Y_j$ for all $j \in \{1, \dots, n\}$ such that $|\frac{j}{n} - x| \leq h_n$. We obtain the following uniform rates of convergence.

Theorem 2.1 *Under model (2.1) and assumptions (F1), (G1), (H1) we have*

$$\sup_{x \in [h_n, 1-h_n]} |\hat{g}(x) - g(x)| = O(h_n^\beta) + O_P\left(\left(\frac{|\log h_n|}{n h_n}\right)^{1/\alpha}\right).$$

Note that the deterministic rate $O(h_n^\beta)$ stems from approximating the regression function by a polynomial, whereas the random rate originates from the observational error. Balancing the two convergence rates by setting $h_n = ((\log n)/n)^{\frac{1}{\alpha\beta+1}}$ gives

$$\sup_{x \in [h_n, 1-h_n]} |\hat{g}(x) - g(x)| = O_P\left(\left(\frac{\log n}{n}\right)^{\frac{\beta}{\alpha\beta+1}}\right). \quad (2.3)$$

This result is of particular interest in the case of irregular error distributions in the sense that large mass is concentrated in a neighborhood of the right endpoint, i.e. $\alpha \in (0, 2)$. Then the rate improves upon the typical optimal rate $O_P(((\log n)/n)^{\frac{\beta}{2\beta+1}})$ for estimating mean regression functions in models with regular errors.

Remark 2.2 Jirak, Meister and Reiß (2014) consider a similar boundary regression estimator while replacing the integral in (2.2) by its Riemann approximation $\sum_{i=1}^n p(\frac{i}{n}) I\{|\frac{i}{n} - x| \leq h_n\}$. For this modified estimator we obtain the same uniform rate of convergence as in Theorem 2.1 by replacing Proposition A.1 in the proof of Theorem 2.1 by Theorem 3.1 in Jirak, Meister and Reiß (2014). ■

Remark 2.3 For Hölder continuous regression functions with exponent $\beta \in (0, 1]$ the estimator reduces to a local maximum, i. e. $\hat{g}(x) = \max\{Y_i \mid i = 1, \dots, n \text{ s. t. } |\frac{i}{n} - x| \leq h_n\}$. In this case we obtain the uniform rate of convergence as given in Theorem 2.1 in the whole design interval. ■

Remark 2.4 Müller and Wefelmeyer (2010) consider a mean regression model $Y_i = m(X_i) + \eta_i$, $i = 1, \dots, n$, with symmetric error distribution supported on $[-a, a]$ (with a unknown). The error distribution function fulfills $F(a - y) \sim 1 - y^\alpha$ for $y \searrow 0$. The local empirical midrange of the responses, i. e.

$$\hat{m}(x) = \frac{1}{2} \left(\min_{\substack{i \in \{1, \dots, n\} \\ |X_i - x| \leq h_n}} Y_i + \max_{\substack{i \in \{1, \dots, n\} \\ |X_i - x| \leq h_n}} Y_i \right)$$

is shown to have pointwise rate of convergence $O(h_n^\beta) + O_P((nh_n)^{-1/\alpha})$ to m if m is Hölder continuous with exponent $\beta \in (0, 1]$. Theorem 2.1 allows us to extend Müller and Wefelmeyer's (2010) results to the general Hölder case ($\beta > 0$) for uniform rates of convergence (in a model with fixed design $X_i = \frac{i}{n}$). To this end we use the mean regression estimator $\hat{m} = (\hat{g} - \hat{\tilde{g}})/2$ with \hat{g} as before and $\hat{\tilde{g}}$ defined analogously, but based on $(\frac{i}{n}, -Y_i)$, $i = 1, \dots, n$. ■

Remark 2.5 In the case $\beta \in (1, 2]$ Hall and Van Keilegom (2009) consider the following local linear boundary regression estimator,

$$\check{g}(x) = \inf \left\{ \alpha_0 \mid (\alpha_0, \alpha_1) \in \mathbb{R}^2 : Y_i \leq \alpha_0 + \alpha_1 \left(\frac{i}{n} - x \right) \forall i \in \{1, \dots, n\} \text{ s. t. } \left| \frac{i}{n} - x \right| \leq h_n \right\}. \quad (2.4)$$

Note that due to $\int_{x-h_n}^{x+h_n} (\alpha_0 + \alpha_1(t-x)) dt = 2\alpha_0 h_n$ this estimator coincides with \hat{g} for $\beta \in (1, 2]$. However, in the case $\beta > 2$ replacing the linear function in (2.4) by a polynomial of order $\lceil \beta \rceil - 1$ renders the estimator \check{g} useless. One obtains $\check{g}(x) = -\infty$ for $x \notin \{\frac{j}{n} \mid j = 1, \dots, n\}$ while $\check{g}(\frac{j}{n}) = Y_j$, $j = 1, \dots, n$. This was already observed by Jirak, Meister and Reiß (2014). ■

Note that the estimator \hat{g} is not smooth. One might prefer to consider a smooth estimator by convoluting \hat{g} with a kernel. Such a modified estimator will also be advantageous when deriving an expansion for the residual based empirical distribution function in the next section. Therefore we define

$$\tilde{g}(x) = \int_{h_n}^{1-h_n} \hat{g}(z) \frac{1}{b_n} K\left(\frac{x-z}{b_n}\right) dz \quad (2.5)$$

and formulate some additional assumptions.

- (K1)** K is a kernel with support $[-1, 1]$ and order $\lfloor \beta \rfloor + 1$, i. e. $\int u^r K(u) du = 0 \forall r = 1, \dots, \lfloor \beta \rfloor$. Further, $K(-1) = K(1) = 0$ and K is differentiable with Lipschitz-continuous derivative K' .

(B1) The sequence $(b_n)_{n \in \mathbb{N}}$ of positive bandwidths satisfies $\lim_{n \rightarrow \infty} b_n = 0$.

(B2) There exists some $0 < \delta < \min(1, \beta - 1)$ such that either

$$\left(h_n^\beta + \left(\frac{|\log h_n|}{nh_n} \right)^{1/\alpha} \right)^{\frac{1}{1+2\delta}} = o(b_n) \text{ and } \delta \leq \frac{\beta - 1}{2}$$

or

$$\left(h_n^\beta + \left(\frac{|\log h_n|}{nh_n} \right)^{1/\alpha} \right)^{\frac{\delta}{3\delta - (\beta - 1)(1 - \delta)}} = o(b_n) \text{ and } \delta > \frac{\beta - 1}{2}.$$

The estimator \tilde{g} is differentiable and we obtain the following uniform rates of convergence for \tilde{g} and its derivative \tilde{g}' .

Theorem 2.6 *Let model (2.1) and the assumptions (F1), (G1) with $\beta \in (1, \infty)$, (H1), (K1), and (B1) be fulfilled. Then for $I_n = [h_n + b_n, 1 - h_n - b_n]$ it holds that*

$$(i) \sup_{x \in I_n} |\tilde{g}(x) - g(x)| = O(b_n^\beta) + O(h_n^\beta) + O_P\left(\left(\frac{|\log h_n|}{nh_n}\right)^{\frac{1}{\alpha}}\right)$$

$$(ii) \sup_{x \in I_n} |\tilde{g}'(x) - g'(x)| = O(b_n^{\beta-1}) + O(b_n^{-1}h_n^\beta) + O_P\left(b_n^{-1}\left(\frac{|\log h_n|}{nh_n}\right)^{\frac{1}{\alpha}}\right) = o_P(1),$$

where the last equality holds under the additional assumption (B2)

$$(iii) \sup_{x, y \in I_n, x \neq y} \frac{|\tilde{g}'(x) - g'(x) - \tilde{g}'(y) + g'(y)|}{|x - y|^\delta} = o_P(1) \text{ under the additional assumption (B2).}$$

3 The error distribution

3.1 Estimation

In this section we consider estimators for the error distribution in model (2.1). For the asymptotic analysis we need an additional assumption on the error distribution.

(F2) The error distribution F is Hölder continuous of order $\alpha \wedge 1$.

We build residuals $\hat{\varepsilon}_i = Y_i - \hat{g}(\frac{i}{n})$, and define the following modified empirical distribution function,

$$\hat{F}_n(y) = \frac{1}{m_n} \sum_{i=1}^n I\{\hat{\varepsilon}_i \leq y\} I\{h_n < \frac{i}{n} \leq 1 - h_n\},$$

where $m_n = \#\{i \in \{1, \dots, n\} \mid h_n < \frac{i}{n} \leq 1 - h_n\} = n - \lfloor nh_n \rfloor - \lceil nh_n \rceil$. We first treat a simple case where the influence of the regression estimation on the residual empirical process is negligible. To this end let F_n denote the standard empirical distribution function of the unobservable errors $\varepsilon_1, \dots, \varepsilon_n$. Then we have the following asymptotic result.

Theorem 3.1 *Let model (2.1) with assumptions $(\mathbf{F1})$, $(\mathbf{G1})$, and $(\mathbf{F2})$ be fulfilled. Let further $\frac{1}{\beta} < \alpha < 2 - \frac{1}{\beta}$ and choose $h_n \sim ((\log n)/n)^{1/(\alpha\beta+1)}$. Then we have*

$$\sup_{y \in \mathbb{R}} |\hat{F}_n(y) - F_n(y)| = o_P(n^{-1/2}).$$

Thus the process $\sqrt{n}(\hat{F}_n - F)$ converges weakly to a centered Gaussian process with covariance function $(y_1, y_2) \mapsto F(y_1 \wedge y_2) - F(y_1)F(y_2)$.

The proof applies the uniform convergence rate from Theorem 2.1 and is given in the appendix. Note that the condition $\frac{1}{\beta} < \alpha < 2 - \frac{1}{\beta}$ is fulfilled for all cases of irregularity $\alpha \in (0, 2)$ subject to a sufficiently smooth regression function. It is needed in the proof of Theorem 3.1 in order to assure that for the rate obtained in (2.3), say a_n , one has

$$|F(y + a_n) - F(y)| = O(a_n^{\alpha \wedge 1}) = o(n^{-1/2}).$$

Remark 3.2 From Theorem 3.1 it follows that under the condition $\alpha \in (1/\beta, 2 - 1/\beta)$ the estimation of the regression function has no impact on the estimation of the irregular error distribution. This is remarkably different from corresponding results on error distribution estimation in mean regression models with regular error distributions. Here the empirical distribution function of residuals, say \check{F}_n , is not asymptotically \sqrt{n} -equivalent to the empirical distribution function of true errors, and $\sqrt{n}(\check{F}_n - F)$ converges to a Gaussian process with complicated covariance structure, compare to Theorem 1 in Akritas and Van Keilegom (2001). In the simple case of a mean regression model with equidistant design and an error distribution F with bounded density f one has

$$\sqrt{n}(\check{F}_n(y) - F_n(y)) = \frac{f(y)}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i + o_P(1)$$

uniformly with respect to $y \in \mathbb{R}$ when the regression function is estimated by a local polynomial estimator, under appropriate bandwidth conditions (see Proposition 3 in Neumeyer and Van Keilegom (2009)). ■

In order to obtain asymptotic results for error distribution estimators beyond the simple case $\frac{1}{\beta} < \alpha < 2 - \frac{1}{\beta}$ a finer analysis is needed. We will use the smooth regression estimator \tilde{g} defined in (2.5) in what follows. Let \tilde{F}_n denote the empirical distribution function based on residuals $\tilde{\varepsilon}_j = Y_j - \tilde{g}(\frac{j}{n})$, i. e.

$$\tilde{F}_n(y) = \frac{1}{m_n} \sum_{j=1}^n I\{\tilde{\varepsilon}_j \leq y\} I\{\frac{j}{n} \in I_n\}$$

where $I_n = [h_n + b_n, 1 - h_n - b_n]$ and $m_n = \#\{j \in \{1, \dots, n\} \mid h_n + b_n \leq \frac{j}{n} \leq 1 - h_n - b_n\} = n - 2\lceil n(h_n + b_n) \rceil + 1$. Then the following asymptotic expansion is valid.

Theorem 3.3 *Let model 2.1 and assumptions **(F1)**, **(G1)**, **(H1)**, **(K1)**, **(B1)**, and **(B2)** be fulfilled. Then*

$$\tilde{F}_n(y) = \frac{1}{n} \sum_{j=1}^n I\{\tilde{\varepsilon}_j \leq y\} + \frac{1}{m_n} \sum_{j=1}^n (F(y + (\tilde{g} - g)(\frac{j}{n})) - F(y)) I\{\frac{j}{n} \in I_n\} + o_P\left(\frac{1}{\sqrt{n}}\right) \quad (2.6)$$

*uniformly with respect to $y \in \mathbb{R}$, if **(F2)** holds and δ in assumption **(B2)** is chosen such that $\delta > \frac{1}{\alpha} - 1$.*

For each $\kappa < 0$ the expansion for $\tilde{F}_n(y)$ holds uniformly with respect to $y \in (-\infty, \kappa]$ without the latter additional conditions on F and δ if F has a bounded density on $(-\infty, \kappa]$.

In order for the expansion to be valid on the whole real line we need for $\beta \in (1, 2]$ that $\frac{1}{\alpha} - 1 < \delta < \beta - 1$ (see assumption **(B2)**), which implies $\alpha > \frac{1}{\beta}$. For $\beta \geq 2$ one obtains the only constraint $\alpha > \frac{1}{2}$.

Next we examine under which conditions the additional term in (2.6) depending on the estimation error is asymptotically negligible. We focus on those arguments y which are bounded away from 0, because in this setting weaker conditions on α and β are needed. Moreover, for the analysis of the tail behavior of the error distribution at 0, tail empirical processes are better suited.

Note that the estimator \hat{g} tends to underestimate the true function because it is defined via a polynomial which is minimal under the constraint that it lies above all observations $(i/n, Y_i)$, which in turn all lie below the true boundary function. As this systematic underestimation does not vanish from (local or global) averaging, we first have to introduce a bias correction.

Let $E_{g \equiv 0}$ denote the expectation if the true regression function is identical 0. For the remaining part of this section, we assume that $E_{g \equiv 0}(\hat{g}(1/2))$ is known or that it can be estimated sufficiently accurately. For example, if the empirical process of residuals shall be used to test a simple null hypothesis, then one may calculate or simulate this expectation under the given null distribution. We define a bias corrected version of the smoothed estimator by

$$\tilde{g}_n^*(x) := \tilde{g}(x) - E_{g \equiv 0}(\hat{g}(1/2)),$$

for $x \in I_n$. The following lemma ensures that the above results for \tilde{g} carry over to this variant if the following condition on the lower tail of F holds:

(F3) There exists $\tau > 0$ such that $F(-t) = o(t^{-\tau})$ as $t \rightarrow -\infty$.

Lemma 3.4 *If model 2.1 holds with g identical 0 and the conditions **(F1)**, **(F3)**, **(G1)**, and **(H1)** are fulfilled, then for all $x \in [h_n, 1 - h_n]$*

$$E(|\hat{g}_n(x)|) = E(|\hat{g}_n(1/2)|) = O\left(\left(\frac{\log n}{nh_n}\right)^{1/\alpha}\right).$$

We need some additional conditions on the rates at which the bandwidths h_n and b_n tend to 0:

$$(H2) \quad h_n = o\left(\min\left(n^{-1/(2\beta)}, n^{-1/(\alpha\beta+1)}\right)\right), \quad n^{\alpha/4-1} \log n = o(h_n)$$

$$(B3) \quad b_n = o\left(\min\left(\left(\frac{nh_n}{\log n}\right)^{2/\alpha} n^{-1}, (nh_n)^{-1/(\alpha\beta)}\right)\right)$$

In particular, these assumptions ensure that the bias terms of order $h_n^\beta + b_n^\beta$ are of smaller order than $n^{-1/2}$ and $(nh_n)^{-1/\alpha}$ and hence asymptotically negligible, and that quadratic terms in the estimation error are uniformly negligible, that is, $\|\tilde{g}_n^* - g\|_\infty^2 = o_P(n^{-1/2})$.

Theorem 3.5 *Suppose model 2.1 holds and assumptions (F1), (F3), (G1), (H1), (H2), (K1), (B1), (B2), and (B3) and F has a bounded density on $(-\infty, \kappa]$ for some $\kappa < 0$. Then*

$$\sup_{y \in (-\infty, \kappa]} \left| \frac{1}{m_n} \sum_{j=1}^n \left(F\left(y + (\tilde{g}_n^* - g)\left(\frac{j}{n}\right)\right) - F(y) \right) I\left\{\frac{j}{n} \in I_n\right\} \right| = o_P(n^{-1/2}).$$

The conditions on h_n and b_n used in Theorem 3.5 can be fulfilled if and only if $\alpha < 3 - 3/(2\beta)$. In particular, this Theorem is applicable if $\beta > 3/2$ and the error distribution is irregular, i.e., $\alpha < 2$. Then the empirical process of the residuals (restricted to $(-\infty, \kappa]$) is asymptotically equivalent to the empirical process of the errors.

3.2 Goodness-of-fit testing

Let $\mathcal{F} = \{F_\vartheta \mid \vartheta \in \Theta\}$ denote a parametric class of error distributions such that for each $\vartheta \in \Theta$, $F_\vartheta(y) = 1 - c_\vartheta |y|^{\alpha_\vartheta} + r_\vartheta(y)$ with $r_\vartheta(y) = o(|y|^{\alpha_\vartheta})$ for $y \nearrow 0$. Our aim is to test the null hypothesis $H_0 : F \in \mathcal{F}$. We assume that $\alpha_\vartheta \in (1/\beta, 2 - 1/\beta)$ for all $\vartheta \in \Theta$, such that under H_0 Theorem 3.1 is valid. Let $\hat{\vartheta}$ denote an estimator for ϑ based on residuals $\hat{\varepsilon}_i = Y_i - \hat{g}(\frac{i}{n})$, $i = 1, \dots, n$. The goodness-of-fit test is based on the empirical process

$$S_n(y) = \sqrt{n}(\hat{F}_n(y) - F_{\hat{\vartheta}}(y)), \quad y \in \mathbb{R}.$$

Under any fixed alternative \hat{g} still consistently estimates g such that \hat{F}_n estimates the error distribution F . If $\hat{\vartheta}$ even under the alternative converges to some $\vartheta^* \in \Theta$, then a consistent hypothesis test can be obtained by rejecting H_0 for large values of, e.g., a Kolmogorov-Smirnov test statistic $\sup_{y \in \mathbb{R}} |S_n(y)|$. Note that under H_0 from Theorem 3.1 it follows that

$$S_n(y) = \sqrt{n}(F_n(y) - F(y)) - \sqrt{n}(F_{\hat{\vartheta}}(y) - F_\vartheta(y)) + o_P(1),$$

where ϑ denotes the true parameter. We consider some examples.

Example 3.6 Consider the mean regression model $Y_i = m(\frac{i}{n}) + \eta_i$, $i = 1, \dots, n$, with symmetric error distribution F as in Remark 2.4. Assume $\beta > 1$ and let \hat{m} be defined as in the aforementioned Remark. Our aim is to test the null hypothesis $H_0 : F \in \mathcal{F} = \{F_\vartheta \mid \vartheta \in \Theta\}$, where F_ϑ denote the distribution function of the uniform distribution on $[-\vartheta, \vartheta]$ (with $\alpha_\vartheta = 1$ for all $\vartheta > 0$). Let residuals be defined as $\hat{\eta}_i = Y_i - \hat{m}(\frac{i}{n})$, $i = 1, \dots, n$, and let

$$\hat{\vartheta}_n = \max \left(\max_{nh_n \leq i \leq n-nh_n} \hat{\eta}_i, -\min_{nh_n \leq i \leq n-nh_n} \hat{\eta}_i \right) = \max_{nh_n \leq i \leq n-nh_n} |\hat{\eta}_i|.$$

Then $|\hat{\vartheta}_n - \vartheta|$ can be bounded by $|\max\{|\eta_i| : nh_n \leq i \leq n-nh_n\} - \vartheta| + \sup_{x \in [h_n, 1-h_n]} |\hat{m}(x) - m(x)|$ and thus is of rate $o_P(n^{-1/2})$. From the definition $F_\vartheta(y) = \frac{x+\vartheta}{2\vartheta} I_{[-\vartheta, \vartheta]}(y) + I_{(\vartheta, \infty)}(y)$ one straightforwardly obtains $F_{\hat{\vartheta}_n}(y) - F_\vartheta(y) = o_P(n^{-1/2})$ uniformly with respect to $y \in \mathbb{R}$. Thus the process S_n converges weakly to a Brownian Bridge B composed with F . The Kolmogorov-Smirnov test statistic $\sup_{y \in \mathbb{R}} |S_n(y)|$ converges in distribution to $\sup_{t \in [0, 1]} |B(t)|$. Thus although our testing problem requires the estimation of a nonparametric function and we have a composite null hypothesis, the same asymptotic distribution appears as in the Kolmogorov-Smirnov test for the simple hypothesis $H_0 : F = F_0$ based on an iid sample with distribution F . ■

Example 3.7 Regard the regression model $Y_i = g(\frac{i}{n}) + \varepsilon_i$, $i = 1, \dots, n$ with regression function g Hölder of order $\beta > 1$ and error distribution function F . We want to test the null hypothesis $H_0 : F \in \mathcal{F} = \{F_\vartheta \mid \vartheta \in \Theta\}$, where $F_\vartheta(y) = e^{\vartheta y} I_{(-\infty, 0)}(y) + I_{[0, \infty)}(y)$ denotes the distribution function of the mirrored exponential distribution. Then $F_\vartheta(y) = 1 - \vartheta|y| + o(|y|)$ for $y \nearrow 0$, i. e. $\alpha_\vartheta = 1$ for all $\vartheta \in \mathbb{R}_{>0}$.

Define $\hat{\vartheta}_n = -(\frac{1}{m_n} \sum_{j=1}^n \hat{\varepsilon}_j I\{h_n < \frac{j}{n} \leq 1 - h_n\})^{-1}$ and note that

$$\begin{aligned} & \hat{\vartheta}_n - \vartheta \\ &= \hat{\vartheta}_n \vartheta \left(\frac{1}{m_n} \sum_{j=1}^n (\varepsilon_j + \frac{1}{\vartheta}) I\{h_n < \frac{j}{n} \leq 1 - h_n\} + \frac{1}{m_n} \sum_{j=1}^n (g(\frac{j}{n}) - \hat{g}(\frac{j}{n})) I\{h_n < \frac{j}{n} \leq 1 - h_n\} \right) \end{aligned}$$

With the central limit theorem and Theorem 2.1 one can deduce

$$\hat{\vartheta}_n - \vartheta = \vartheta^2 \frac{1}{m_n} \sum_{j=1}^n (\varepsilon_j + \frac{1}{\vartheta}) I\{h_n < \frac{j}{n} \leq 1 - h_n\} + o_{P_\vartheta} \left(\frac{1}{\sqrt{n}} \right)$$

and thus for $y \in (-\infty, 0)$ and some $\xi_{n,\vartheta}$ between $\hat{\vartheta}_n$ and ϑ ,

$$\begin{aligned} F_{\hat{\vartheta}_n}(y) &= e^{\hat{\vartheta}_n y} = e^{\vartheta y} + e^{\vartheta y} y (\hat{\vartheta}_n - \vartheta) + \frac{1}{2} e^{\xi_{n,\vartheta} y} y^2 (\hat{\vartheta}_n - \vartheta)^2 \\ &= F_\vartheta(y) + e^{\vartheta y} y \vartheta^2 \frac{1}{m_n} \sum_{j=1}^n (\varepsilon_j + \frac{1}{\vartheta}) I\{h_n < \frac{j}{n} \leq 1 - h_n\} + o_{P_\vartheta} \left(\frac{1}{\sqrt{n}} \right) \end{aligned}$$

uniformly in y . Now analogously to the proof of Theorem 19.23 in van der Vaart (2000) we can deduce weak convergence of

$$S_n(y) = \sqrt{n}(F_n(y) - F(y)) - e^{\vartheta y} \vartheta^2 \frac{\sqrt{n}}{m_n} \sum_{j=1}^n (\varepsilon_j + \frac{1}{\vartheta}) I\{h_n < \frac{j}{n} \leq 1 - h_n\} + o_{P_\vartheta}(1), y \in \mathbb{R},$$

to a Gaussian process with covariance function $(y_1, y_2) \mapsto F_\vartheta(y_1 \wedge y_2) - F_\vartheta(y_1)F_\vartheta(y_2) - e^{\vartheta(y_1+y_2)} y_1 y_2 \vartheta^2$, where the covariance function follows by simple calculations and the fact that $E_\vartheta[I\{\varepsilon_1 \leq y\}(\varepsilon_1 + \frac{1}{\vartheta})] = y e^{\vartheta y}$. The asymptotic distribution of the Cramér-von-Mises test statistic $\int S_n(y)^2 dF_{\hat{\vartheta}_n}(y)$ is distribution free and we can use tabled quantiles for the test, see e. g. Stephens (1976). ■

3.3 Simulations

To study the small sample performance of our goodness-of-fit test we investigate its behaviour on simulated data according to Example 3.6. The observations are generated with $g(x) = 0.5 \sin(2\pi x) + 4x$ and errors ε_i that are distributed according to the density function $f_\zeta(y) = 0.5(1 - |y|)^\zeta$ for different values of $\zeta > -1$. Note that for $\zeta = 0$ the null hypothesis $H_0 : \varepsilon_i \sim U[-\vartheta, \vartheta]$ holds whereas for $\zeta \neq 0$ the test should reject the null hypothesis. In Figure 1 the results for different sample sizes and level 5% are displayed. They are based on 200 Monte Carlo repetitions in each case and calculated with the Cramér-von-Mises test statistic. The bandwidth is chosen as $0.6n^{-\frac{1}{3}}$ (this is the optimal bandwidth for $\alpha = 1$ and $\beta = 2$ apart from the log-term) and the regression function is estimated by the local linear approach. It can be seen that the asymptotic level is approximated well and the power increases with increasing ζ as well as with increasing n .

To understand the influence of the bandwidth we simulate the same model with $h_n = c \cdot n^{-\frac{1}{3}}$ for different c from $c = 0.2$ to $c = 1.2$. The results are similar to those displayed in Figure 1 for c between 0.2 and 0.8 but for ca. $c \geq 0.9$ the power and the size get worse especially for the small sample sizes $n = 50$ and $n = 100$.

A Appendix: Proofs

A.1 Auxiliary results

The following proposition can be verified by an obvious modification of the proof of Theorem 3.1 by Jirak, Meister and Reiß (2014).

Proposition A.1 *Assume that model (2.1) holds and that the regression function g fulfills condition **(G1)** for some $\beta \in (0, \beta^*]$ and some $c_g \in [0, c^*]$. Then there exist constants $L_{\beta^*, c^*}, L_{\beta^*} > 0$ and a natural number j_{β^*} (depending only on the respective subscripts) such*

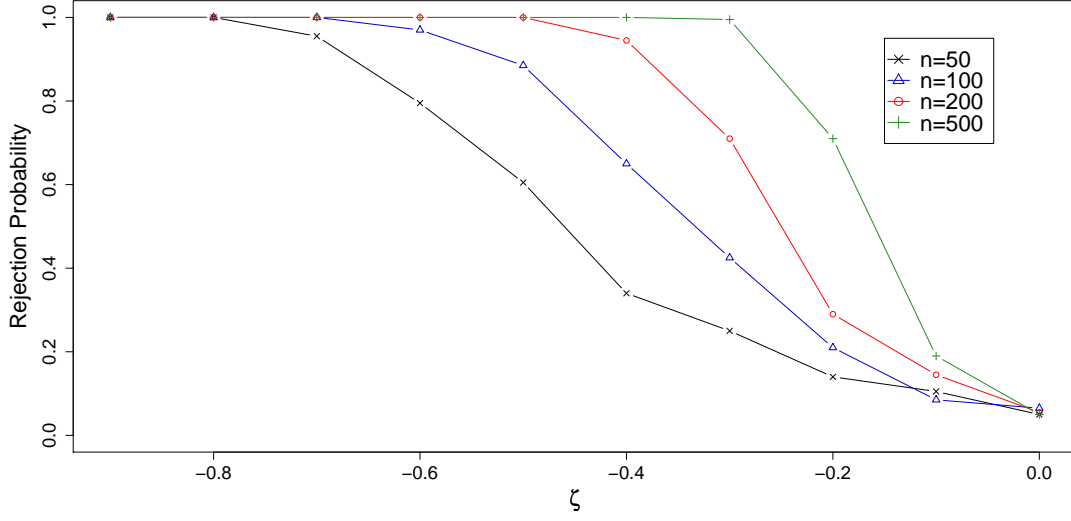


Figure 1: Monte-Carlo simulations for Example 3.6

that

$$|\hat{g}(x) - g(x)| \leq L_{\beta^*, c^*} h_n^\beta + L_{\beta^*} \max_{1 \leq j \leq 2j_{\beta^*}} \left(\min_{i: -1+(j-1)/j_{\beta^*} \leq |i/n-x|/h_n \leq -1+j/j_{\beta^*}} |\varepsilon_i| \right).$$

Lemma A.2 *Under assumptions (F1) and (H1) for any fixed set I_1, \dots, I_m of disjoint non-degenerate subintervals of $[-1, 1]$ we have*

$$\sup_{x \in [h_n, 1-h_n]} \max_{1 \leq j \leq m} \min_{\substack{i \in \{1, \dots, n\} \\ (i/n-x)/h_n \in I_j}} (-\varepsilon_i) = O_P\left(\left(\frac{|\log h_n|}{nh_n}\right)^{1/\alpha}\right).$$

Proof. Let $b_n := (|\log h_n|/(nh_n))^{1/\alpha}$. Obviously it suffices to prove that for all non-degenerate subintervals $I \subset [-1, 1]$ there exists a constant L such that

$$\lim_{n \rightarrow \infty} P\left\{ \sup_{x \in [h_n, 1-h_n]} \min_{\substack{i \in \{1, \dots, n\} \\ (i/n-x)/h_n \in I}} |\varepsilon_i| > Lb_n \right\} = 0.$$

Denote by $d = \sup I - \inf I > 0$ the diameter of I and let $d_n := \lfloor nh_n d \rfloor - 1$ and $l_n := \lfloor n/d_n \rfloor$. Then for all $x > 0$

$$\begin{aligned} P\left\{ \sup_{x \in [h_n, 1-h_n]} \min_{\substack{i \in \{1, \dots, n\} \\ (i/n-x)/h_n \in I}} |\varepsilon_i| > x \right\} &\leq P\left\{ \max_{j \in \{1, \dots, n-d_n\}} \min_{i \in \{j, \dots, j+d_n\}} |\varepsilon_i| > x \right\} \\ &\leq P\left\{ \max_{\substack{l \in \{0, \dots, l_n\} \\ l \text{ even}}} M_{n,l} > x \right\} + P\left\{ \max_{\substack{l \in \{0, \dots, l_n\} \\ l \text{ odd}}} M_{n,l} > x \right\} \end{aligned}$$

with

$$M_{n,l} := \max_{j \in \{ld_n+1, \dots, (l+1)d_n\}} \min_{i \in \{j, \dots, j+d_n\}} |\varepsilon_i|.$$

Since the random variables $M_{n,l}$ for l even are iid, we have

$$P\left\{\max_{\substack{l \in \{0, \dots, l_n\} \\ l \text{ even}}} M_{n,l} > x\right\} = 1 - (1 - P\{M_{n,0} > x\})^{\lfloor l_n/2 \rfloor + 1},$$

and an analogous equation holds for the maxima over the odd numbered block maxima $M_{n,l}$.

Let G be the cdf of $|\varepsilon_i|$. If $M_{n,0}$ exceeds x , then there is a smallest index $j \in \{1, \dots, d_n\}$ for which $\min_{i \in \{j, \dots, j+d_n\}} |\varepsilon_i| > x$. Hence

$$\begin{aligned} P\{M_{n,0} > x\} &= P\left\{\min_{i \in \{1, \dots, 1+d_n\}} |\varepsilon_i| > x\right\} + \sum_{j=2}^{d_n} P\left\{|\varepsilon_{j-1}| \leq x, \min_{i \in \{j, \dots, j+d_n\}} |\varepsilon_i| > x\right\} \\ &= (1 - G(x))^{d_n+1} + (d_n - 1)G(x)(1 - G(x))^{d_n+1} \\ &\leq (1 + d_n G(x))(1 - G(x))^{d_n}. \end{aligned}$$

To sum up, we have shown that

$$P\left\{\sup_{x \in [h-n, 1-h_n]} \min_{\substack{i \in \{1, \dots, n\} \\ (i/n-x)/h_n \in I}} |\varepsilon_i| > Lb_n\right\} \leq 2 \left(1 - \left(1 - (1 + d_n G(Lb_n))(1 - G(Lb_n))^{d_n}\right)^{\lfloor l_n/2 \rfloor + 1}\right)$$

It remains to show that the right hand side tends to 0 for sufficiently large L which is true if and only if

$$(1 + d_n G(Lb_n))(1 - G(Lb_n))^{d_n} = o(1/l_n).$$

This is an immediate consequence of $1/l_n \sim dh_n$ and

$$\begin{aligned} G(Lb_n) &= cL^\alpha \frac{|\log h_n|}{nh_n} (1 + o(1)) \\ \implies (1 - G(Lb_n))^{d_n} &= \exp\left(-nh_n dcL^\alpha \frac{|\log h_n|}{nh_n} (1 + o(1))\right) \\ \implies (1 + d_n G(Lb_n))(1 - G(Lb_n))^{d_n} &= O\left(|\log h_n| \exp(-cdL^\alpha |\log h_n| (1 + o(1)))\right) = o(h_n) \end{aligned}$$

if $cdL^\alpha > 1$.

□

A.2 Proof of Theorem 2.1

The assertion directly follows from Proposition A.1 and Lemma A.2.

□

A.3 Proof of Theorem 2.6

(i) One obtains straightforwardly

$$\sup_{x \in I_n} |\tilde{g}(x) - g(x)| \leq \sup_{x \in I_n} \left| \int_{h_n}^{1-h_n} (\hat{g}(z) - g(z)) \frac{1}{b_n} K\left(\frac{x-z}{b_n}\right) dz \right|$$

$$\begin{aligned}
& + \sup_{x \in I_n} \left| \int_{h_n}^{1-h_n} (g(z) - g(x)) \frac{1}{b_n} K \left(\frac{x-z}{b_n} \right) dz \right| \\
& \leq \sup_{z \in [h_n, 1-h_n]} |\hat{g}(z) - g(z)| O(1) + \sup_{x \in I_n} \left| \int_{-1}^1 (g(x - ub_n) - g(x)) K(u) du \right| \\
& \leq O(h_n^\beta) + O_P \left(\left(\frac{|\log h_n|}{nh_n} \right)^{\frac{1}{\alpha}} \right) \\
& + b_n^{\lfloor \beta \rfloor} \sup_{x \in I_n} \left| \frac{1}{\lfloor \beta \rfloor!} \int_{-1}^1 u^{\lfloor \beta \rfloor} (g^{(\lfloor \beta \rfloor)}(x - \tau ub_n) - g^{(\lfloor \beta \rfloor)}(x)) K(u) du \right|
\end{aligned}$$

with application of Theorem 2.1, a Taylor expansion of g of order $\lfloor \beta \rfloor$ and assumption **(K1)**. Now an application of assumption **(G1)** on the last term yields the desired result.

(ii) Since g is bounded on $[h_n, 1 - h_n]$ and $\sup_{x \in [h_n, 1-h_n]} |\hat{g}(x) - g(x)| = o_P(1)$, we can suppose that \hat{g} is bounded on $[h_n, 1 - h_n]$ too. From the equicontinuity of K' it follows that $\hat{g}(z)K'((x-z)/b_n)/b_n^2$ is equicontinuous in x for $z \in [h_n, 1 - h_n]$ and we can exchange integration and differentiation and obtain

$$\sup_{x \in I_n} |\tilde{g}'(x) - g'(x)| = \sup_{x \in I_n} \left| \int_{h_n}^{1-h_n} \hat{g}(z) \frac{1}{b_n^2} K' \left(\frac{x-z}{b_n} \right) dz - g'(x) \right|.$$

Integration by parts yields

$$\int_{h_n}^{1-h_n} g(z) \frac{1}{b_n^2} K' \left(\frac{x-z}{b_n} \right) dz = \int_{h_n}^{1-h_n} g'(z) \frac{1}{b_n} K \left(\frac{x-z}{b_n} \right) dz$$

since $K(-1) = K(1) = 0$ and therefore

$$\begin{aligned}
\sup_{x \in I_n} |\tilde{g}'(x) - g'(x)| & \leq \sup_{x \in I_n} \left| \int_{h_n}^{1-h_n} (\hat{g}(z) - g(z)) \frac{1}{b_n^2} K' \left(\frac{x-z}{b_n} \right) dz \right| \\
& + \sup_{x \in I_n} \left| \int_{h_n}^{1-h_n} (g'(z) - g'(x)) \frac{1}{b_n} K \left(\frac{x-z}{b_n} \right) dz \right| \\
& \leq \sup_{z \in [h_n, 1-h_n]} |\hat{g}(z) - g(z)| O\left(\frac{1}{b_n}\right) + \sup_{x \in I_n} \left| \int_{-1}^1 (g'(x - ub_n) - g'(x)) K(u) du \right|
\end{aligned}$$

The assertion follows by Theorem 2.1, a Taylor expansion of g' of order $\lfloor \beta \rfloor - 1$ and the assumptions **(K1)** and **(G1)**.

(iii) We distinguish between the cases $|x - y| > a_n$ and $|x - y| \leq a_n$ for some suitable sequence $(a_n)_{n \in \mathbb{N}}$ with $\lim_{n \rightarrow \infty} a_n = 0$ discussed later. For the first case we obtain

$$\begin{aligned}
& \sup_{x, y \in I_n, |x-y| > a_n} \frac{|\tilde{g}'(x) - g'(x) - \tilde{g}'(y) + g'(y)|}{|x - y|^\delta} \\
& \leq 2 \sup_{x \in I_n} |\tilde{g}'(x) - g'(x)| a_n^{-\delta}
\end{aligned}$$

$$= \left(O(b_n^{\beta-1}) + \left(O(h_n^\beta) + O_P \left(\left(\frac{|\log h_n|}{nh_n} \right)^{\frac{1}{\alpha}} \right) \right) \frac{1}{b_n} \right) a_n^{-\delta}. \quad (2.7)$$

For the second case we use a decomposition like in the proof of (ii)

$$\begin{aligned} & \sup_{x,y \in I_n, 0 < |x-y| \leq a_n} \frac{|\tilde{g}'(x) - g'(x) - \tilde{g}'(y) + g'(y)|}{|x-y|^\delta} \\ & \leq \sup_{x,y \in I_n, 0 < |x-y| \leq a_n} \frac{\left| \int_{h_n}^{1-h_n} (\hat{g}(z) - g(z)) \frac{1}{b_n^2} \left(K' \left(\frac{x-z}{b_n} \right) - K' \left(\frac{y-z}{b_n} \right) \right) dz \right|}{|x-y|^\delta} \\ & \quad + \sup_{\substack{x,y \in I_n \\ 0 < |x-y| \leq a_n}} \frac{|g'(x) - g'(y)|}{|x-y|^\delta} + \sup_{\substack{x,y \in I_n \\ 0 < |x-y| \leq a_n}} \frac{\left| \int_{h_n}^{1-h_n} g'(z) \frac{1}{b_n} \left(K \left(\frac{x-z}{b_n} \right) - K \left(\frac{y-z}{b_n} \right) \right) dz \right|}{|x-y|^\delta}. \end{aligned}$$

By Lipschitz continuity of K' and Theorem 2.1 the first term on the right hand side is of rate

$$\left(O(h_n^\beta) + O_P \left(\left(\frac{|\log h_n|}{nh_n} \right)^{\frac{1}{\alpha}} \right) \right) \frac{1}{b_n^3} O(a_n^{1-\delta}). \quad (2.8)$$

For the second term one obtains the rate $a_n^{1-\delta} = o(1)$ by differentiability of g' in the case $\beta \geq 2$. In the case $\beta \in (1, 2)$ by assumption **(G1)** for the same term one obtains the rate $a_n^{\beta-1-\delta}$ which is $o(1)$ by assumption **(B2)**.

The last term on the right hand side can be reformulated as

$$\sup_{\substack{x,y \in I_n \\ 0 < |x-y| \leq a_n}} \frac{\left| \int_{-1}^1 (g'(x - h_n u) - g'(y - h_n u)) K(u) du \right|}{|x-y|^\delta}$$

and is thus of the same order as the second term by assumption **(G1)**.

To finish the proof one needs a sequence $a_n = o(1)$ such that the remaining rates (2.7) and (2.8) are of negligible order $o_P(1)$. Using the notation $\vartheta_n = h_n^\beta + (|\log h_n|/(nh_n))^{1/\alpha}$ the existence of a sequence $(a_n)_{n \in \mathbb{N}}$ such that

$$b_n^{\beta-1} + \frac{\vartheta_n}{b_n} = o(a_n^\delta) \text{ and } a_n^{1-\delta} = o\left(\frac{b_n^3}{\vartheta_n}\right)$$

is equivalent to

$$b_n^{\beta-1} + \frac{\vartheta_n}{b_n} = o\left(\left(\frac{b_n^3}{\vartheta_n}\right)^{\frac{\delta}{1-\delta}}\right)$$

which in turn is equivalent to the bandwidth condition in assumption **(B2)**. \square

A.4 Proof of Theorem 3.1

By choice of the bandwidth sequence $(h_n)_n$ equation (2.3) holds. Let $\tilde{a}_n = ((\log n)/n)^{\beta/(\alpha\beta+1)}$, then, by assumption, $\tilde{a}_n = o(n^{-1/(2(\alpha\wedge 1))})$. From this and (2.3) one can deduce the existence of some sequence $(a_n)_n$ with

$$a_n = o(n^{-\frac{1}{2(\alpha\wedge 1)}}) \quad (2.9)$$

such that

$$P\left(\sup_{x \in [h_n, 1-h_n]} |\hat{g}(x) - g(x)| \leq a_n\right) \xrightarrow{n \rightarrow \infty} 1.$$

Since

$$\hat{F}_n(y) = \frac{1}{m_n} \sum_{j=1}^n I\{\varepsilon_j \leq y + (\hat{g} - g)(\frac{j}{n})\} I\{h_n < \frac{j}{n} \leq 1 - h_n\}$$

this implies that with probability converging to one

$$\sqrt{n}(\bar{F}_n(y - a_n) - F_n(y)) \leq \sqrt{n}(\hat{F}_n(y) - F_n(y)) \leq \sqrt{n}(\bar{F}_n(y + a_n) - F_n(y)) \quad \forall y \in \mathbb{R},$$

where we define $\bar{F}_n(y) := \frac{1}{m_n} \sum_{j=1}^n I\{\varepsilon_j \leq y\} I\{h_n < \frac{j}{n} \leq 1 - h_n\}$. We take a closer look at the bounds. To this end let $E_n(y, s) = \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor ns \rfloor} (I\{\varepsilon_j \leq y\} - F(y))$, then the sequential empirical process $(E_n(y, s))_{y \in \mathbb{R}, s \in [0, 1]}$ converges weakly to a Kiefer process, see e. g. Theorem 2.12.1 in van der Vaart and Wellner (1996). Now uniformly with respect to $y \in \mathbb{R}$ (since $a_n \searrow 0$)

$$\begin{aligned} & \sqrt{n}(\bar{F}_n(y \pm a_n) - F_n(y)) \\ &= \frac{n}{m_n} \left(E_n(y \pm a_n, 1 - h_n) - E_n(y, 1 - h_n) - E_n(y \pm a_n, h_n) + E_n(y, h_n) \right) \\ & \quad + \sqrt{n}(F(y \pm a_n) - F(y)) + \sqrt{n}(\bar{F}_n(y) - F_n(y)) \\ &= o_P(1) + \sqrt{n}(\bar{F}_n(y) - F_n(y)) \end{aligned}$$

since $\frac{n}{m_n} = O(1)$, the process E_n is asymptotically equicontinuous, assumption **(F2)** and (2.9) It remains to show that $\sqrt{n}(\bar{F}_n(y) - F_n(y)) = o_P(1)$ uniformly in $y \in \mathbb{R}$. It holds that

$$\begin{aligned} \bar{F}_n(y) - F_n(y) &= \frac{1}{m_n} \sum_{j=1}^n (I\{\varepsilon_j \leq y\} - F(y)) I\{h_n < \frac{j}{n} \leq 1 - h_n\} - \frac{1}{n} \sum_{j=1}^n (I\{\varepsilon_j \leq y\} - F(y)) \\ &= \left(\frac{1}{m_n} - \frac{1}{n}\right) \sqrt{n} \left(E_n(y, 1 - h_n) - E_n(y, h_n) \right) \\ & \quad - \frac{1}{\sqrt{n}} \left(E_n(y, h_n) + E_n(y, 1) - E_n(y, 1 - h_n) \right) \\ &= O\left(\frac{h_n}{\sqrt{n}}\right) (E_n(y, 1) + o_P(1)) + o_P\left(\frac{1}{\sqrt{n}}\right) \end{aligned}$$

because $h_n \rightarrow 0$, $E_n(y, 0) \equiv 0$ and the process E_n is asymptotically equicontinuous. Since the limit of the standard empirical process $(E_n(y, 1))_{y \in \mathbb{R}}$ is tight we obtain $\bar{F}_n(y) - F_n(y) = o_P(n^{-1/2})$ uniformly with respect to $y \in \mathbb{R}$ and, hence, the assertion of the Theorem. \square

A.5 Proof of Theorem 3.3

Define for any interval $I \subset \mathbb{R}$ and constant $k > 0$ the following class of differentiable functions,

$$C_k^{1+\delta}(I) = \left\{ d : I \rightarrow \mathbb{R} \mid \max\left\{\sup_{x \in I} |d(x)|, \sup_{x \in I} |d'(x)|\right\} + \sup_{x, y \in I, x \neq y} \frac{|d'(x) - d'(y)|}{|x - y|^\delta} \leq k \right\}.$$

Then with Theorem 2.6 (i) - (iii) we obtain $P((\tilde{g} - g) \in C_{1/2}^{1+\delta}(I_n)) \rightarrow 1$ for $n \rightarrow \infty$. From this we can deduce the existence of some sequence of random functions $d_n : [0, 1] \rightarrow \mathbb{R}$ with $d_n(x) = (\tilde{g} - g)(x)$ for all $x \in I_n$ and $P(d_n \in C_1^{1+\delta}([0, 1])) \rightarrow 1$ for $n \rightarrow \infty$.

To prove the asserted expansion we are going to apply Theorem 2.11.9 in van der Vaart and Wellner (1996) to the process

$$G_n(\varphi) = \sum_{j=1}^n (Z_{nj}(\varphi) - E[Z_{nj}(\varphi)]), \quad \varphi = (y, d) \in \mathcal{F},$$

where $\mathcal{F} = \{(y, d) \mid y \in \mathbb{R}, d \in C_1^{1+\delta}([0, 1])\}$, equipped with the semimetric ρ defined by

$$\rho((y, d), (y', d')) = \max\left\{\sup_{x \in [0, 1]} \sup_{\gamma \in C_1^{1+\delta}([0, 1])} |F(y + \gamma(x)) - F(y' + \gamma(x))|, \sup_{x \in [0, 1]} |d(x) - d'(x)|\right\},$$

and

$$Z_{nj}(\varphi) = \frac{\sqrt{n}}{m_n} I\{\varepsilon_j \leq y + d(\frac{j}{n})\} I\{\frac{j}{n} \in I_n\} - \frac{1}{\sqrt{n}} I\{\varepsilon_j \leq y\}.$$

We will show at the end of the proof that (\mathcal{F}, ρ) is a totally bounded semimetric space. In the following we check the assumptions in the Theorem 2.11.9 in van der Vaart and Wellner (1996). The proof is analogous to the proof of Lemma 3 in Neumeyer and Van Keilegom (2009) (see the online supporting information to that article). The first condition holds since $\sup_{\varphi \in \mathcal{F}} |Z_{nj}(\varphi)| \leq \frac{\sqrt{n}}{m_n} + \frac{1}{\sqrt{n}} = o(1)$ and hence $\forall \eta > 0 \exists n_0 \in \mathbb{N}$ s. t. for all $n \geq n_0$

$$\sum_{j=1}^n E \left[\sup_{\varphi \in \mathcal{F}} |Z_{nj}(\varphi)| I\left\{\sup_{\varphi \in \mathcal{F}} |Z_{nj}(\varphi)| > \eta\right\} \right] \leq n \left(\frac{\sqrt{n}}{m_n} + \frac{1}{\sqrt{n}} \right) I \left\{ \frac{\sqrt{n}}{m_n} + \frac{1}{\sqrt{n}} > \eta \right\} = 0.$$

Further note that

$$\begin{aligned} n(Z_{nj}(y, d) - Z_{nj}(y', d'))^2 &= \left(\frac{n}{m_n} (I\{\varepsilon_j \leq y + d(\frac{j}{n})\} - I\{\varepsilon_j \leq y' + d(\frac{j}{n})\}) I\{\frac{j}{n} \in I_n\} \right. \\ &\quad \left. - (I\{\varepsilon_j \leq y\} - I\{\varepsilon_j \leq y'\}) \right. \\ &\quad \left. + \frac{n}{m_n} (I\{\varepsilon_j \leq y' + d(\frac{j}{n})\} - I\{\varepsilon_j \leq y' + d'(\frac{j}{n})\}) I\{\frac{j}{n} \in I_n\} \right)^2. \end{aligned}$$

Now with $\varphi = (y, d)$ and $\varphi' = (y', d')$ we have

$$\sup_{\varphi, \varphi' \in \mathcal{F}, \rho(\varphi, \varphi') < \varepsilon_n} \sum_{j=1}^n E \left[(Z_{nj}(\varphi) - Z_{nj}(\varphi'))^2 \right]$$

$$\begin{aligned}
&\leq \frac{4}{m_n} \sup_{\varphi, \varphi' \in \mathcal{F}, \rho(\varphi, \varphi') < \bar{\epsilon}_n} \sum_{j=1}^n \frac{n}{m_n} E \left[\left(I\{\varepsilon_j \leq y + d(\frac{j}{n})\} - I\{\varepsilon_j \leq y' + d(\frac{j}{n})\} \right)^2 \right] I\{\frac{j}{n} \in I_n\} \\
&\quad + \frac{4}{n} \sup_{\varphi, \varphi' \in \mathcal{F}, \rho(\varphi, \varphi') < \bar{\epsilon}_n} \sum_{j=1}^n E \left[\left(I\{\varepsilon_j \leq y\} - I\{\varepsilon_j \leq y'\} \right)^2 \right] \\
&\quad + \frac{4}{m_n} \sup_{\varphi, \varphi' \in \mathcal{F}, \rho(\varphi, \varphi') < \bar{\epsilon}_n} \sum_{j=1}^n \frac{n}{m_n} E \left[\left(I\{\varepsilon_j \leq y' + d(\frac{j}{n})\} - I\{\varepsilon_j \leq y' + d'(\frac{j}{n})\} \right)^2 \right] I\{\frac{j}{n} \in I_n\} \\
&= \frac{4}{m_n} \sup_{\varphi, \varphi' \in \mathcal{F}, \rho(\varphi, \varphi') < \bar{\epsilon}_n} \sum_{j=1}^n \frac{n}{m_n} \left| F(y + d(\frac{j}{n})) - F(y' + d(\frac{j}{n})) \right| I\{\frac{j}{n} \in I_n\} \\
&\quad + 4 \sup_{\varphi, \varphi' \in \mathcal{F}, \rho(\varphi, \varphi') < \bar{\epsilon}_n} |F(y) - F(y')| \\
&\quad + \frac{4}{m_n} \sup_{\varphi, \varphi' \in \mathcal{F}, \rho(\varphi, \varphi') < \bar{\epsilon}_n} \sum_{j=1}^n \frac{n}{m_n} \left| F(y' + d(\frac{j}{n})) - F(y' + d'(\frac{j}{n})) \right| I\{\frac{j}{n} \in I_n\} \\
&\leq 4 \frac{n}{m_n} \sup_{\varphi, \varphi' \in \mathcal{F}, \rho(\varphi, \varphi') < \bar{\epsilon}_n} \sup_{x \in [0,1]} \sup_{d \in C_1^{1+\delta}([0,1])} |F(y + d(x)) - F(y' + d(x))| \\
&\quad + 4 \sup_{\varphi, \varphi' \in \mathcal{F}, \rho(\varphi, \varphi') < \bar{\epsilon}_n} |F(y) - F(y')| \\
&\quad + 4 \frac{n}{m_n} \sup_{\varphi, \varphi' \in \mathcal{F}, \rho(\varphi, \varphi') < \bar{\epsilon}_n} \sup_{x \in [0,1]} |F(y' + d(x)) - F(y' + d'(x))| \\
&= O(\bar{\epsilon}_n) + O(\bar{\epsilon}_n^{\alpha \wedge 1}) = o(1)
\end{aligned}$$

by the definition of ρ since $\frac{n}{m_n} = O(1)$ and by assumption **(F2)**. This is the second condition in the aforementioned Theorem.

The validity of the third assumption can be proved in the same way as in the proof of Lemma 3 in Neumeyer and Van Keilegom (2009) (see their online supporting information). We just have to replace one step since we do not assume a bounded error density. By assumption **(F2)** it holds that

$$\sup_{y \in \mathbb{R}} \sup_{x \in [0,1]} |F(y + d_m^U(x)) - F(y + d_m^L(x))| \leq \sup_{x \in [0,1]} c_F |d_m^U(x) - d_m^L(x)|^{\alpha \wedge 1} \quad (2.10)$$

where $c_F < \infty$ is the Hölder constant of F . Therefore with brackets that fulfill $\sup_{x \in [0,1]} |d_m^U(x) - d_m^L(x)| \leq \eta^{2(\alpha^{-1} \vee 1)}$ we get

$$\sup_{y \in \mathbb{R}} \sup_{x \in [0,1]} |F(y + d_m^U(x)) - F(y + d_m^L(x))| = O(\eta^2).$$

With this the bracketing number $N_{[]}(\eta, \mathcal{F}, L_n^2)$ is of the order

$$O \left(\eta^{-2} \exp \left(\kappa \eta^{-\frac{2(\alpha^{-1} \vee 1)}{1+\delta}} \right) \right) \quad (2.11)$$

and therefore, if $\delta > \frac{1}{\alpha} - 1$,

$$\int_0^{\bar{\epsilon}_n} \sqrt{\log(N_{[]}(\eta, \mathcal{F}, L_n^2))} d\eta = O \left(\int_0^{\bar{\epsilon}_n} \sqrt{\log(\eta^{-1})} d\eta \right) + O \left(\int_0^{\bar{\epsilon}_n} \sqrt{\eta^{-\frac{2(\alpha^{-1} \vee 1)}{1+\delta}}} d\eta \right)$$

$$\xrightarrow[n \rightarrow \infty]{} 0 \quad \text{for } \bar{\epsilon}_n \searrow 0,$$

which is the third condition in Theorem 2.11.9 in van der Vaart and Wellner (1996).

Now weak convergence of the process $(G_n(\varphi))_{\varphi \in \mathcal{F}}$ follows from its marginal convergence, which can be shown easily using Cramér-Wold's device and Lindeberg's condition. Therefore $(G_n(\varphi))_{\varphi \in \mathcal{F}}$ is also asymptotically equicontinuous, that is

$$\sup_{\varphi, \varphi' \in \mathcal{F}, \rho(\varphi, \varphi') < \bar{\epsilon}_n} |G_n(\varphi) - G_n(\varphi')| = o_P(1)$$

for all $\bar{\epsilon}_n \searrow 0$. With $\varphi = (y, d_n)$ and $\varphi' = (y, 0)$ we get

$$\begin{aligned} \sup_{y \in \mathbb{R}} \left| \sum_{j=1}^n \left(\frac{\sqrt{n}}{m_n} (I\{\varepsilon_j \leq y + d_n(\frac{j}{n})\} - F(y + d_n(\frac{j}{n}))) I\{\frac{j}{n} \in I_n\} - \frac{1}{\sqrt{n}} (I\{\varepsilon_j \leq y\} - F(y)) \right) \right. \\ \left. - \sum_{j=1}^n \left(\frac{\sqrt{n}}{m_n} (I\{\varepsilon_j \leq y\} - F(y)) I\{\frac{j}{n} \in I_n\} - \frac{1}{\sqrt{n}} (I\{\varepsilon_j \leq y\} - F(y)) \right) \right| = o_P(1) \end{aligned}$$

which together with

$$\left| \sum_{j=1}^n \left(\frac{\sqrt{n}}{m_n} (I\{\varepsilon_j \leq y\} - F(y)) I\{\frac{j}{n} \in I_n\} - \frac{1}{\sqrt{n}} (I\{\varepsilon_j \leq y\} - F(y)) \right) \right| = o_P(1)$$

and the fact that $d_n(x) = (\tilde{g} - g)(x) \forall x \in I_n$ implies our assertion.

It remains to show that (\mathcal{F}, ρ) is a totally bounded semimetric space. This can be proved in the same way as in the proof of Lemma 3 in Neumeyer and Van Keilegom (2009) with one small change: We again replace the bounds $|F(y) - F(z)| \leq \sup_{x \in \mathbb{R}} |f(x)| |y - z|$ by $|F(y) - F(z)| \leq c_F |y - z|^{\alpha \wedge 1}$. This may change the bracketing number $N_{[]}(\eta, \mathcal{F}, \rho)$, but it is still finite for all $\eta > 0$, which implies that \mathcal{F} is totally bounded.

Note that if we restrict to $y \in (-\infty, \kappa]$ for some $\kappa < 0$ and assume a bounded error density on $(-\infty, \kappa]$, instead of the inequality (2.10) one applies the mean value theorem. Consequently the bracketing number in (2.11) changes to $O(\eta^{-2} \exp(\kappa \eta^{-2/(1+\delta)}))$ and for finiteness of the bracketing integral no additional assumption on $\delta \in (0, 1)$ is needed. \square

In the remaining proofs, we use the index n for the estimators to emphasis the dependence on the sample size and to distinguish between estimators and polynomials corresponding to a given sample on the one hand and corresponding objects in a limiting setting on the other hand.

A.6 Proof of Lemma 3.4

Proposition A.1 and the proof of Lemma A.2 show that there exist constants $c, d > 0$ depending only on β and c_g such that $E(\hat{g}_n(x)) \leq cE(M_{n,0})$ and $P\{M_{n,0} > t\} \leq (1 + dnh_n(1 - F(-t)))(F(-t))^{dnh_n}$ for all $t > 0$.

Let $a_n := a(\log n / (nh_n))^{1/\alpha}$ for a suitable constant $a > 0$ and fix some $t_0 > 0$ such that $(1 - F(-t)) / (c_F t^\alpha) \in (1/2, 2)$ for all $t \in (0, t_0]$. Then

$$\begin{aligned} E(M_{n,0}) &= \int_0^\infty P\{M_{n,0} > t\} dt \\ &\leq a_n + \int_{a_n}^{t_0} (1 + dnh_n(1 - F(-t))) (F(-t))^{dnh_n} dt + (1 + dnh_n) \int_{t_0}^\infty ((F(-t))^{dnh_n} dt. \end{aligned}$$

Now

$$\begin{aligned} &\int_{a_n}^{t_0} (1 + dnh_n(1 - F(-t))) (F(-t))^{dnh_n} dt \\ &\leq \int_{a_n}^{t_0} (1 + 2c_F dnh_n t^\alpha) \left(1 - \frac{c_F}{2} t^\alpha\right)^{dnh_n} dt \\ &\leq (1 + 2c_F d) nh_n \int_{a_n}^{t_0} t^\alpha \exp\left(-\frac{c_F}{2} dnh_n t^\alpha\right) dt \\ &\leq (1 + 2c_F d) nh_n \frac{t_0}{\alpha} \int_{a_n^\alpha}^{t_0^\alpha} \exp\left(-\frac{c_F}{2} dnh_n u\right) du \\ &\leq (1 + 2c_F d) nh_n \frac{t_0}{\alpha} \exp\left(-\frac{c_F}{2} da^\alpha \log n\right) \\ &= o(n^{-\xi}) \end{aligned}$$

for all $\xi > 0$ if a is chosen sufficiently large. Hence the assertion follows from

$$\int_{t_0}^\infty (F(-t))^{dnh_n} dt \leq nh_n ((F(-t_0))^{dnh_n} + \int_{nh_n}^\infty t^{-dnh_n} dt = o(n^{-\xi})$$

for all $\xi > 0$. □

A.7 Proof of Theorem 3.5

As the density f is bounded and Lipschitz continuous, one has

$$\begin{aligned} |F(y + (\tilde{g}_n^* - g)(j/n)) - F(y) - f(y)(\tilde{g}_n^* - g)(j/n)| &= \left| \int_0^{(\tilde{g}_n^* - g)(j/n)} f(y+t) - f(y) dt \right| \\ &= O((\tilde{g}_n^* - g)^2(j/n)) \\ &= O\left(\left(\frac{\log n}{nh_n}\right)^{1/\alpha}\right) \end{aligned}$$

uniformly for $y \leq y_0$. Hence the remainder term can be approximated by a sum of estimation errors as follows:

$$\left| \frac{1}{m_n} \sum_{j=1}^n (F(y + (\tilde{g}_n^* - g)(j/n)) - F(y)) I\{j/n \in I_n\} - \frac{f(y)}{m_n} \sum_{j=1}^n (\tilde{g}_n^* - g)(j/n) I\{j/n \in I_n\} \right|$$

$$= O\left(\frac{1}{m_n} \sum_{j=1}^n (\tilde{g}_n^* - g)^2(j/n) I\{j/n \in I_n\}\right) = O_P\left(h_n^\beta + b_n^\beta + \left(\frac{\log n}{nh_n}\right)^{2/\alpha}\right) = o_P(n^{-1/2})$$

where for the last conclusions we have used Theorem 2.1, Lemma 3.4 and the assumptions **(H2)** and **(B3)**. Thus the assertion follows if we show that

$$\frac{1}{m_n} \sum_{j=1}^n (\tilde{g}_n^* - g)(j/n) I\{j/n \in I_n\} = o_P(n^{-1/2}).$$

To this end, note that $\tilde{g}_n^*(x)$ and $\tilde{g}_n^*(y)$ are independent for $|x - y| > 2(h_n + b_n)$. For simplicity, we assume that $2n(h_n + b_n) =: k_n$ is a natural number. If we split the whole sum into blocks with k_n consecutive summands, then all blocks with odd numbers are independent and all blocks with even numbers are independent. It suffices to show that

$$\begin{aligned} \frac{1}{m_n} \sum_{\ell=1}^{\lfloor n/(2k_n) \rfloor} \Delta_{n,2\ell-1} &= o_P(n^{-1/2}) \\ \frac{1}{m_n} \sum_{\ell=1}^{\lfloor n/(2k_n) \rfloor} \Delta_{n,2\ell} &= o_P(n^{-1/2}) \end{aligned}$$

where $\Delta_{n,\ell} = \sum_{j=(\ell-1)k_n+1}^{\ell k_n} (\tilde{g}_n^* - g)(j/n)$, $1 \leq \ell \leq \lfloor n/k_n \rfloor$. We only consider the second sum, because the first convergence obviously follows by the same arguments.

It suffices to verify the following two conditions.

$$E(\Delta_{n,2\ell}^2) = o(k_n) \tag{2.12}$$

$$E(\Delta_{n,2\ell}) = o(n^{-1/2}k_n) = o(n^{1/2}(h_n + b_n)) \tag{2.13}$$

uniformly for all $1 \leq \ell \leq \lfloor n/(2k_n) \rfloor$. For then

$$E\left(\sum_{\ell=1}^{\lfloor n/(2k_n) \rfloor} \Delta_{n,2\ell}\right)^2 = \sum_{\ell=1}^{\lfloor n/(2k_n) \rfloor} \text{Var}(\Delta_{n,2\ell}) + \left(\sum_{\ell=1}^{\lfloor n/(2k_n) \rfloor} E\Delta_{n,2\ell}\right)^2 = o(n)$$

which implies the assertion.

To prove (2.12), note that according to Lemma 2.12, Proposition A.1, and the proofs of Lemma A.2 and of Theorem 2.6(i), there exist constants $c_1, c_2, c_3 > 0$ (depending only on β , c_g and the kernel K) such that

$$\sup_{x \in I_n} |\tilde{g}_n^*(x) - g(x)| \leq c_1(h_n^\beta + b_n^\beta + \max(M_1^*, M_2^*))$$

where M_1^*, M_2^* are independent random variables such that $P\{M_i^* > t\} \leq 1 - (1 - P\{M_{n,0} > t\})^{c_2(h_n + b_n)/h_n}$ with

$$P\{M_{n,0} > t\} \leq (1 + c_3nh_n(1 - F(-t)))(F(-t))^{c_3nh_n}.$$

Because $k_n(h_n^\beta + b_n^\beta) = o(k_n^{1/2})$ by **(H2)** and **(B3)**, it suffices to show that

$$E((M_i^*)^2) = \int_0^\infty P\{M_i^* > t^{1/2}\} dt = o(1/k_n). \quad (2.14)$$

Fix some $t_0 \in (0, (2c_F)^{-2/\alpha})$ such that $(1 - F(-t))/(c_F t^\alpha) \in (1/2, 2)$ for all $t \in (0, t_0]$. In what follows, c denotes a generic constant (depending only on β, c_g, c_F and K) which may vary from line to line. Applying the inequality $(1 - u)^\rho \leq \exp(-\rho u)$, which holds for all $\rho > 0$ and $u \in (0, 1)$, twice, we obtain for $(nh_n/\log n)^{-2/\alpha} < t \leq t_0$

$$\begin{aligned} P\{M_i^* > t^{1/2}\} &\leq 1 - [1 - (1 + c_3 nh_n 2c_F t^{\alpha/2})(1 - c_F t^{\alpha/2}/2)^{c_3 nh_n}]^{c_2(h_n + b_n)/h_n} \\ &\leq 1 - \exp\left(-cn(h_n + b_n)t^{\alpha/2} \exp(-c_3 c_F nh_n t^{\alpha/2}/2)\right) \\ &\leq cn(h_n + b_n)t^{\alpha/2} \exp(-c_3 c_F nh_n t^{\alpha/2}/2). \end{aligned}$$

Therefore, for sufficiently large $d > 0$,

$$\begin{aligned} &\int_0^{t_0^2} P\{M_i^* > t^{1/2}\} dt \\ &\leq d\left(\frac{nh_n}{\log n}\right)^{-2/\alpha} + ct_0 n(h_n + b_n) \int_{d(nh_n/\log n)^{-2/\alpha}}^{t_0} t^{\alpha/2-1} \exp(-c_3 c_F nh_n t^{\alpha/2}/2) dt \\ &\leq o(1/(n(h_n + b_n))) + ct_0 n(h_n + b_n) \exp(-c_3 c_F d^{\alpha/2} \log n) \\ &= o(1/(n(h_n + b_n))) \end{aligned} \quad (2.15)$$

where in the last but one step we apply the conditions **(B3)** and **(H2)**. Now, assertion (2.14) (and hence (2.12)) follows from

$$\begin{aligned} \int_{t_0^2}^\infty P\{M_i^* > t^{1/2}\} dt &\leq \int_{t_0^2}^\infty 1 - [1 - c_3 nh_n (F(-t^{1/2}))^{c_3 nh_n}]^{c_2(h_n + b_n)/h_n} dt \\ &\leq \int_{t_0^2}^\infty 1 - \exp\left(-cn(h_n + b_n)(F(-t^{1/2}))^{c_3 nh_n}\right) dt \\ &\leq cn(h_n + b_n) \left(nh_n (F(-t_0))^{c_3 nh_n} + \int_{nh_n}^\infty t^{-\tau c_3 nh_n} dt\right) \\ &= o(n^{-\xi}) \end{aligned}$$

for all $\xi > 0$ where we have used **(H2)** and **(F3)**.

To establish (2.13), first note that for a kernel K of order $d + 1$ with $d := \lfloor \beta \rfloor$

$$\begin{aligned} E(\tilde{g}_n(x) - g(x)) &= E\left(\int_{-1}^1 \left(\hat{g}_n(x + b_n u) - \sum_{j=0}^d \frac{g^{(j)}(x)}{j!} (b_n u)^j\right) K(u) du\right) \\ &= \int_{-1}^1 E(\hat{g}_n(x + b_n u) - g(x + b_n u)) K(u) du + O(b_n^\beta) \end{aligned}$$

uniformly for all $x \in [h_n + b_n, 1 - h_n - b_n]$. In view of **(H2)** and **(B3)**, it thus suffices to show that

$$|E(\hat{g}_n(x) - g(x)) - E_{g=0}(\hat{g}_n(1/2))| = o(n^{-1/2}) \quad (2.16)$$

uniformly for Lebesgue almost all $x \in [h_n, 1 - h_n]$.

Recall that $\hat{g}_n(x) = \tilde{p}_n(0)$ where \tilde{p}_n is a polynomial of degree d on $[-1, 1]$ that solves the linear optimization problem

$$\int_{-1}^1 \tilde{p}_n(t) dt \rightarrow \min!$$

under the constraints

$$\tilde{p}_n\left(\frac{i/n - x}{h_n}\right) \geq Y_i, \quad \forall i \in [n(x - h_n), n(x + h_n)].$$

Define polynomials

$$q_x(t) := \sum_{k=0}^d \frac{1}{k!} g^{(k)}(x) (h_n t)^k, \quad p_n(t) := (nh_n)^{1/\alpha} (\tilde{p}_n(t) - q_x(t)), \quad t \in [-1, 1].$$

Then $q_x((u - x)/h_n)$ is the Taylor expansion of order d of $g(u)$ at x and the estimation error can be written as

$$\hat{g}_n(x) - g(x) = (nh_n)^{-1/\alpha} p_n(0). \quad (2.17)$$

Note that p_n is a polynomial of degree d that solves the linear optimization problem

$$\int_{-1}^1 p_n(t) dt \rightarrow \min!$$

subject to

$$p_n\left(\frac{i/n - x}{h_n}\right) \geq (nh_n)^{1/\alpha} \bar{\varepsilon}_i, \quad \forall i \in [n(x - h_n), n(x + h_n)], \quad (2.18)$$

with

$$\bar{\varepsilon}_i := \varepsilon_i + g(i/n) - q_x\left(\frac{i/n - x}{h_n}\right).$$

We now use point process techniques to analyze the asymptotic behavior of this linear program.

Denote by

$$N_n := \sum_{i \in [n(x - h_n), n(x + h_n)]} \delta_{((i/n - x)/h_n, (nh_n)^{1/\alpha} \bar{\varepsilon}_i)}$$

a point process of standardized error random variables. Then the constraints (2.18) can be reformulated as $N_n(A_{p_n}) = 0$ where $A_f := \{(t, u) \in [-1, 1] \times \mathbb{R} \mid u > f(t)\}$ denotes the open epigraph of a function f .

It is well known (see, e.g., Resnick (2007), Theorem 6.3) that N_n converges weakly to a Poisson process N on $[-1, 1] \times \mathbb{R}$ with intensity measure $2c_F U_{[-1, 1]} \otimes \nu_\alpha$ where ν_α has Lebesgue

density $x \mapsto \alpha|x|^{\alpha-1}I(-\infty, 0)$, because by **(H2)** $|\bar{\varepsilon}_i - \varepsilon_i| = g(i_n) - q_x((i/n - x)/h_n) = O(h_n^\beta) = o((nh_n)^{-1/\alpha})$ uniformly for all $i \in [n(x - h_n), n(x + h_n)]$

$$E(N_n([-1, 1] \times (-1, \infty))) = 2nh_n P\{\bar{\varepsilon}_1 > (nh_n)^{-1/\alpha}\} \rightarrow 2c_F.$$

By Skorohod's representation theorem, we may assume that the convergence holds a.s.

Next we analyze the corresponding linear program in the limiting model to minimize $\int_{-1}^1 p(t) dt$ over polynomials of degree d subject to $N(A_p) = 0$. In what follows we use a representation of the Poisson process as $N = \sum_{i=1}^\infty \delta_{(T_i, Z_i)}$ where T_i are independent random variables which are uniformly distributed on $[-1, 1]$. First we prove by contradiction that the optimal solution is almost surely unique. Suppose that there exist more than one solution. From the theory of linear programs it is known that then there exists a solution p such that $J := \{j \in \mathbb{N} \mid p(T_j) = Z_j\}$ has at most d elements. Because p is bounded and N has a.s. finitely many points in any bounded set, $\eta := \inf\{|p(T_i) - Z_i| \mid i \in \mathbb{N} \setminus J\} > 0$ a.s. Since p is an optimal solution, for all polynomials Δ of degree d satisfying $\Delta(T_j) = 0$, $j \in J$, and $\|\Delta\|_\infty < \eta$ must satisfy $\int_{-1}^1 \Delta(t) dt = 0$, because both $p + \Delta$ and $p - \Delta$ satisfy the constraints $N(A_{p \pm \Delta}) = 0$. In particular, for all sufficiently small $\tau > 0$ and all polynomials q of degree $d - |J|$, $\Delta(t) = \tau \prod_{i \in J} (t - T_i) q(t)$ is of that type. Write $\prod_{i \in J} (t - T_i)$ in the form $t^{|J|} + \sum_{l=0}^{|J|-1} a_l t^l$. Then necessarily

$$\int_{-1}^1 \prod_{i \in J} (t - T_i) t^j dt = \frac{2}{|J| + j + 1} I\{|J| + j \text{ even}\} + \sum_{l=0}^{|J|-1} \frac{2a_l}{l + j + 1} I\{l + j \text{ even}\} = 0,$$

for all $j \in \{0, \dots, d - |J|\}$. This implies that $(T_i)_{i \in J}$ lies on a manifold $M_{|J|, d}$ of dimension $|J| - (d - |J| + 1) = 2|J| - d - 1$ which only depends on $|J|$ and d . However, by Proposition A.1, $\|p\|_\infty \leq K_d Z_{\max}$ where

$$Z_{\max} := \max_{1 \leq i \leq j_d} \min\{|Z_i| \mid T_i \in [-1 + (j - 1)/j_d, -1 + j/j_d]\}.$$

The above conclusion contradicts $P\{Z_{\max} > K\} \rightarrow 0$ as $K \rightarrow \infty$, since

$$P\{\exists J \subset \mathbb{N} : |J| \leq d, (T_j)_{j \in J} \in M_{|J|, d}, \max_{j \in J} |Z_j| \leq K_d K\} = 0$$

for all $K > 0$ (i.e., the fact that among finitely many values T_i a.s. there does not exist a subset which lies on a given manifold of lower dimension).

Therefore the solution p must be a.s. unique which in turn implies that it is a basic feasible solution, i.e., the $|J| \geq d + 1$. On the other hand, because the intensity measure of N is absolutely continuous, $|J| \leq d + 1$ a.s. and thus $|J| = d + 1$. Because of $N_n \rightarrow N$ a.s., one has $N_n([-1, 1] \times [-K_d Z_{\max}, \infty)) = N([-1, 1] \times [-K_d Z_{\max}, \infty)) =: M$ for sufficiently large n . Moreover, one can find a numeration of the points $(T_{n,i}, Z_{n,i})$, $1 \leq i \leq M$, of N_n and (T_i, Z_i) , $1 \leq i \leq M$, of N in $[-1, 1] \times [-K_d Z_{\max}, \infty)$ such that $(T_{n,i}, Z_{n,i}) \rightarrow (T_i, Z_i)$.

Next we prove that the solution to the linear program to minimize $\int_{-1}^1 p_n(t) dt$ subject to $N_n(A_{p_n}) = 0$ is eventually unique with $p_n \rightarrow p$ a.s. Since any optimal solution can be written as a convex combination of a basic feasible solution, w.l.o.g. we may assume that $J_n := \{1 \leq i \leq M \mid p_n(T_{n,i}) = Z_{n,i}\}$ has at least $d + 1$ elements. The polynomial p_n is uniquely determined by this set J_n . Suppose that along a subsequence n' the set $J_{n'}$ is constant, but not equal to J . Then p'_n converges uniformly to the polynomial \bar{p} of degree d that is uniquely determined by the conditions $\bar{p}(T_i) = Z_i$ for all $i \in J_{n'}$. In particular, \bar{p} is different from the optimal polynomial p for the limit Poisson process, but it satisfies the constraints $N(A_p) = 0$. Thus $\int_{-1}^1 \bar{p}(t) dt > \int_{-1}^1 p(t) dt$. On the other hand, for all $\eta > 0$ the polynomial $p + \eta$ eventually satisfies the constraints $N_n(A_{p+\eta}) = 0$ and thus $\int_{-1}^1 p(t) + \eta dt \geq \int_{-1}^1 \bar{p}_n(t) dt$, which leads to a contradiction.

Hence, $J_n = J$ for all sufficiently large n and the optimal solution p_n for N_n is unique and it converges uniformly to the optimal solution p for the Poisson process N . Moreover, using the relation $(p_n(T_{n,i}))_{i \in J} = (Z_{n,i})_{i \in J}$ (which is a system of linear equation in the coefficients of p_n), $p_n(0)$ can be calculated as $w_n^t(Z_{n,j})_{j \in J}$ for some vector w_n which converges to a limit vector w (corresponding to the analogous relation for p).

Exactly the same arguments apply if we replace $\bar{\varepsilon}_i$ with ε_i , which corresponds to the case that g is identical 0. Since the points $(\tilde{T}_{n,i}, \tilde{Z}_{n,i})$ of the pertaining point process equal $(T_{n,i}, Z_{n,i} - (nh_n)^{1/\alpha}(g(i/n) - q_x((i/n) - x)/h_n))$ and thus $|\tilde{Z}_{n,i} - Z_{n,i}| \leq c_g(nh_n)^{1/\alpha}h_n^\beta$, the difference of the resulting values for optimal polynomial at 0 is bounded by a multiple of $(nh_n)^{1/\alpha}h_n^\beta$. In view of (2.17), we may conclude that the estimation errors differ only by a multiple of $h_n^\beta = o(n^{-1/2})$, which finally yields (2.16) and thus the assertion. \square

References

- Akritis, M. and Van Keilegom, I. (2001). Nonparametric estimation of the residual distribution. *Scand. J. Statist.* **28**, 549–567.
- Gijbels, I., Mammen, E., Park, B. and Simar, L. (2000). On estimation of monotone and concave frontier functions. *J. Amer. Statist. Assoc.* **94**, 220–228.
- Gijbels, I. and Peng, L. (2000). Estimation of a support curve via order statistics. *Extremes* **3**, 251–277.
- Girard, S. and Jacob, P. (2008). Frontier estimation via kernel regression on high power-transformed data. *J. Multivariate Anal.* **99**, 403–420.
- Girard, S., Guillou, A. and Stupfler, G. (2013). Frontier estimation with kernel regression on high order moments. *J. Multivariate Anal.* **116**, 172–189.

- Hall, P. and Van Keilegom, I. (2009). Nonparametric “regression” when errors are positioned at end-points. *Bernoulli* **15**, 614–633.
- Hall, P., Park, B.U. and Stern, S.E. (1998). On polynomial estimators of frontiers and boundaries. *J. Multivariate Anal.* **66**, 71–98.
- Härdle, W., Park, B.U. and Tsybakov, A.B. (1995). Estimation of non-sharp support boundaries. *J. Multivariate Anal.* **55**, 205–218.
- Jirak, M., Meister, A. and Reiß, M. (2014). Adaptive estimation in nonparametric regression with one-sided errors. *Ann. Statist.*, to appear.
available at <http://arxiv.org/abs/1305.6430>
- Meister, A. and Reiß, M. (2013). Asymptotic equivalence for nonparametric regression with non-regular errors. *Probab. Th. Rel. Fields* **155**, 201–229.
- Müller, U.U. and Wefelmeyer W. (2010). Estimation in Nonparametric Regression with Non-Regular Errors. *Comm. Statist. Theory Methods* **39**, 1619–1629.
- Neumeyer, N. and Van Keilegom, I. (2009). Change-Point Tests for the Error Distribution in Nonparametric Regression. *Scand. J. Statist.* **36**, 518–541.
online supporting information available at
<http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9469.2009.00639.x/supinfo>
- Reiß, M. and Selk, L. (2014). Efficient nonparametric functional estimation for one-sided regression. *Preprint*, available at <http://arxiv.org/abs/1407.4229>.
- Resnick, S.I. (2007). *Heavy-Tail Phenomena*. Springer.
- Stephens, M.A. (1976). Asymptotic Results for Goodness-of-Fit Statistics with Unknown Parameters. *Ann. Statist.* **4**, 357–369.
- van der Vaart, A.W. (2000). *Asymptotic Statistics*. Cambridge University Press.
- van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.